Content

What is Neuroinformatics?

Where does it fit in with neuroscience:
experimental, computational and cognitive,
(and clinical)

What's happening in Neuroinformatics
- **INCF**
- **The CARMEN project**

Neuroinformatics and Spikes

Section 1: What is Neuroinformatics?

Informatics applied to Neuroscience (of all sorts)

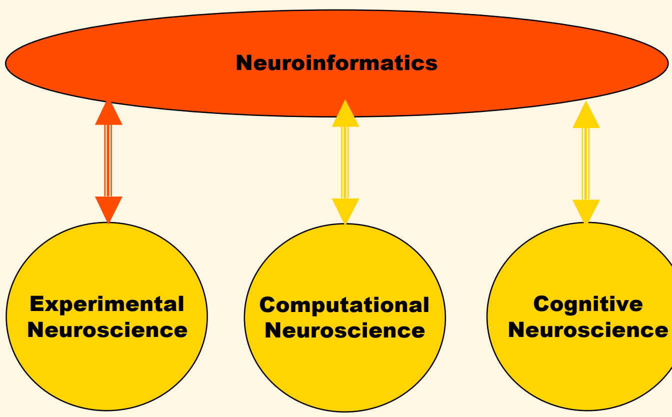
Experimental Neuroscience:
- **Data recording, data analysis have used computers for a long time.**
- **But a great deal more can be achieved**

Cognitive Neuroscience
- **Modelling,**
  - **Matching models to more experimental data**
  - **Matching models to known appropriate behaviour**

Computational Neuroscience
- **Matching models to more experimental data**
- **Defining and running more sophisticated models**
- **Running models in real time**

None of this is entirely new!

Neuroinformatics and Experimental Neuroscience



What is Neuroinformatics bringing to Experimental Neuroscience?

Getting leverage from e-Science capabilities to allow better use of data.

Example 1: Dataset re-use:
Experimenter does experiment, records data, analyses data, writes the paper, perhaps makes the data available to a small number of colleagues.

...and then?

The dataset languishes,
- **first on a spinning disk,**
- **then later on some DVD's,**
- **then later still, is lost to view, as the experimenter changes lab,...**

Yet the data could be of use to other researchers...

?

What are the basic problems holding back dataset re-use? (1)

Two basic problems: Data volume and format, and Metadata

Data access and format problems:
- **Volume**
  - **The data volume is very high: a single experiment can generate 20Gbytes/hour**
- **Media**
  - **"format creep": readers for the physical media become hard to find**
    - **5 inch, 8 inch floppy discs, half inch magnetic tape, old removable discs, old tape cartridges...**

- **Structure**
  - **The data format is a particular structure**
  - **The structure may be**
    - **Proprietary: defined by a particular piece of software, and not made public**
    - **Locally generated: defined by a locally written piece of software, but not necessarily well documented**
    - **Public, but no suitable converter exists for the intending user**

What are the basic problems holding back dataset re-use? (2)

Metadata problems
- **The data itself is useless unless the re-user knows exactly what the data represents.**
- **(Presumably the experimenter knew)**
  - **But did they record this information in an accessible way?**

- **Metadata is data about the dataset**
  - **How was it generated?**
  - **What were the experimental conditions?**
  - **What was the culture, or what preparation, or what animal,…?**
  - **What was the temperature of the recording? Etc. etc.**

- **If the data is to be readily re-used these metadata problems need to be solved in a directly usable way**
  - **Simply describing the protocol in English is not enough**
    - **Can't automate reading J. Neurosci. yet!**
  - **There needs to be an automatically processable way of describing the experimental protocol.**

  - **Particularly true is datasets are used for a large-scale survey of data**
    - **e.g. for data-mining.**

How have other scientific endeavours overcome related problems?

Physics, astronomy, genetic biology all share data very effectively
- **Sharing their datasets has enabled huge strides to be made.**

They have found ways to store large volumes of data in sharable formats.
They have found ways of describing their experimental protocols.

But (say some neuroscientists…)
- **It's easy for them:**
  - **The datasets are all about the same things**
  - **They are all well-defined and agreed physical quantities**
    - **Physical measurements, electromagnetic radiation, sky maps, DNA sequences…**

- **Every preparation is different**
- **Every brain is different, even between individuals in a population**
- **The precise stimuli used are different**

Is Neuroscience Different?

If the differences between neurophysiological datasets are so irreconcilable,
are we really doing science at all, in terms of experiments being repeatable?

If the results from the experiment are useful scientifically,
then re-using the datasets might allow
re-evaluation of the experiment, or
provide evidence for future hypotheses.

Sharing and re-using data may allow bigger strides to be made in understanding neural systems.

Enabling Neuroinformatics based collaboration

Solving data storage and format problems

Data storage: large data stores are now relatively inexpensive.
Even so, multi-terabyte data storage remains expensive (but soluble financially!)

Data format creep:
This is a problem that results from data being archived, then not accessed.
If the data is accessed reasonably often, then the problem will be identified
and data moved on to more recent storage media.

Data structure problems
Force users to adopt a common format?
Alienates users: they won't do it unless they can see real benefits
Support documented formats
Perhaps adopt common internal format, providing translators to & from this format
Rely on proprietary format owners to come aboard because of customer pressure

Enabling Neuroinformatics based collaboration: solving metadata problems

Difficult problem: there are a number of attempts at solving it:

BrainML: (Cornell) brainml.org
"BrainML is a developing initiative to provide a standard XML metaformat for exchanging neuroscience data. It focuses on layered definitions built over a common core in order to support community-driven extension."

NeuroML: http://www.neuroml.org/
"NeuroML is an XML-based description language for defining and exchanging neuronal cell, network and modeling data including reconstructions of cell anatomy, membrane physiology, electrophysiological data, network connectivity, and model specification"

Relevant not only for Neuroinformatics and experimental neuroscience: a cross-cutting problem for Computational and Cognitive neuroscience as well.

Solving Metadata problems continued:

As well as metadata systems for neuronal systems, there are related metadata systems which can be used by BrainML and NeuroML,

ChannelML: for defining ion channel models
MorphML: for defining the morphology of a neuron
SBML: Systems Biology markup language: models of biochemical reaction networks
CellML:to store and exchange computer-based mathematical models

SBML is particularly well advanced: see http://sbml.org/index.psp.

MathML: for describing mathematical notation and capturing both its structure and content. See http://www.w3.org/Math/

Metadata is a big but soluble problem.

It is a multi-level problem, but the systems above provide a multi-level solution.

Enabling Neuroinformatics based collaboration: Sociological problems

There is a reluctance to permit re-use amongst some experimental neuroscientists.

What do experimental neuroscientists get from allowing others to reuse their data?

- **If the answer is only better science, then some experimental neuroscientists will not come on board.**
- **They need to be convinced sharing that their hard-earned datasets will be of benefit to them**
  - **Names on papers?**
  - **The ability to be involved in the further research?**
  - **At the very least, some credit!**

Some neuroscientists fear that their data will be used without their knowledge

- **or worse, that errors in the data (or in their analyses) will show up!**

There is therefore some reticence amongst the experimental neuroscience community.

Solving sociological problems

There are partly technical solutions:

Security aspects on the holding of data:

- **Ensure that datasets can be secured: for example that they can only be re-used with the experimenter's permission.**
- **Security is critically important for holding of data which is still being analysed prior to publication.**

...and there are less technical aspects:

Bringing experimental neuroscientists on board

- **Ensuring that the Neuroinformatics community is properly cross-disciplinary, with good representation from the experimentalists.**

Getting journals on-side

- **Many journals are demanding that raw/processed data be made available in order to check results.**

Neuroinformatics and data interpretation

Leverage from e-Science capabilities to allow easier interpretation of data.

Providing analysis capabilities
Neuroscience data can be analysed in many ways.

For example, for voltage based neuronal cell recordings, there are three basic classes of analyses:

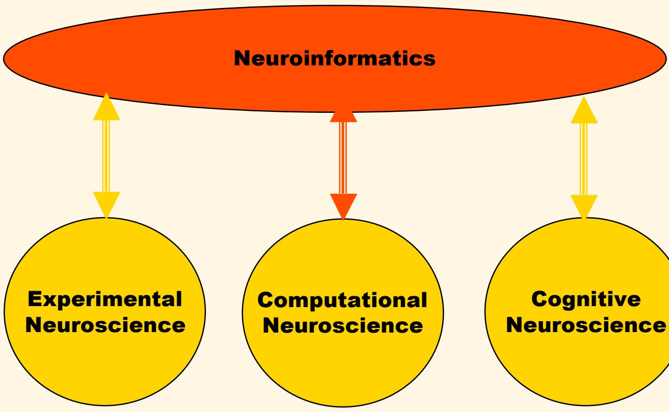
Spike detection:
Finding the action potentials from the recorded analogue signal

Spike sorting
Assigning a particular neuron to each detected spike (where more than one neuron is being picked up by each electrode)

Spike train analysis
Analysing the spike trains from each neuron to attempt to understand the information processing (or the system functionality).

There are many techniques for each of these. Providing services such as these (and others) can make a Neuroinformatics system much more useful.

Neuroinformatics and Computational Neuroscience



What can Neuroinformatics bring to Computational Neuroscience?

Computational neuroscience is has two primary aims:

- **To attempt to understand how neural systems work by building and running models**
- **To attempt to build novel systems by re-creating neural systems which perform these tasks very well**
  - **(Neuro-biologically inspired computing: hardware or software)**

As with experimental neuroscience, many modellers do not share models

- **Lack of time to document the models appropriately**
- **Difficulties in specifying exactly what a model does**
  - **(and there can be sociological issues here as well)**
  - ***Why* should I share my model?**

Neuroinformatics includes

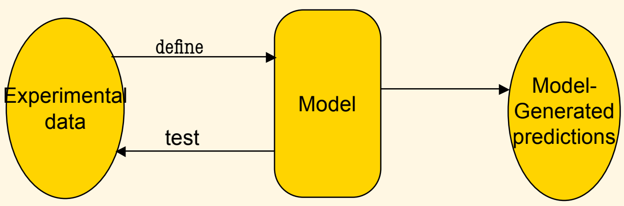
- **Sharing neuroscience data, enabling better-informed models to be built**
- **Techniques for automating the specification of models**

How might sharing neuroscience data help build better models?

All models are based on data.

Very often, they are based on an abstracted understanding of a great deal of real data

- **For example, models of spike time dependent plasticity are not derived from the actual experimental measurements directly**
  - **but from mathematical models extracted from the analyses of the data**
- **More data means better models**



Having more data makes it easier to test models, and possibly to test their predictions as well.

Neuroinformatics and sharing automating the specification of models

What do we mean when we say that we *understand* some neuronal system?

One possibility is that we can build a model…
(which might be abstract or mathematical or computational)
…of the system, and that the model will show the same behaviour as the aspect of the system that we are trying to explain.

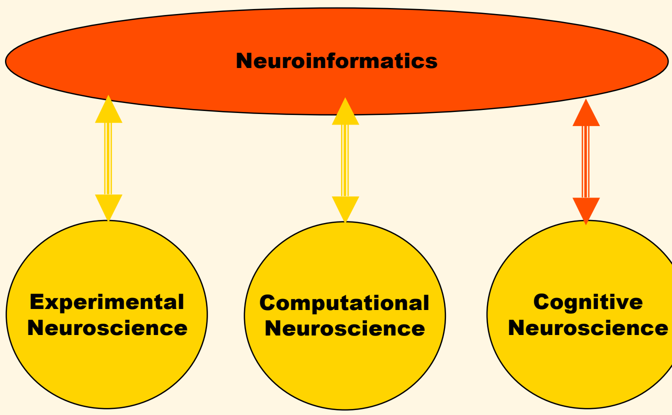
How do we know what a model actually implements?

Is there a specification (in some language) that describes what the model does?
a series of mathematical equations?
some specification language text (in, e.g. MathML and NeuroML)?

Without this, re-using the model is problematic.

One aim of Neuroinformatics is to make specifying and communicating what models do simpler and easier.
Enabling sharing because the specification is precise.

Neuroinformatics and Cognitive Neuroscience



What can Neuroinformatics bring to Cognitive Neuroscience?

Cognitive neuroscience attempts to understand cognitive capabilities in a way compatible with Neuroscience.

It does so primarily by building models …
(of many different types: mathematical, abstract, computational,…)
… which display some of the characteristics of the cognitive system of interest.

Models are at a higher level than those of computational neuroscience
reflecting the higher-level status of cognition, over (plain) computation

The primary difference from computational neuroscience, is that the data being explained are cognitive and behavioural, rather than neural signals.
Sharing neuroscience data does not directly help generate better models
But sharing neuroscience data may help to bridge the gap between cognition and neural signals.

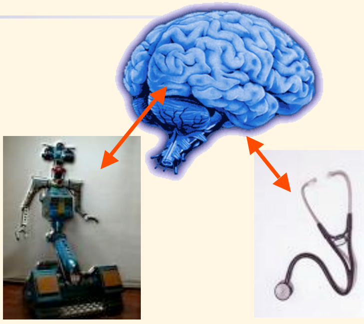
Specifying and communicating models applies just as strongly as it did to computational neuroscience models.
Possibly even more more, as these models sometimes lack rigour!
And NeuroML, MathML etc. can help here

Neuroinformatics and clinical neuroscience

Two sides to understanding neural systems

How does it do what it does
- How can we understand it?
- How can we build machines that use the ideas in the brain

What can be done when it goes wrong?
- Mental illness
- Brain diseases
- Trauma



This meeting is really about the first side
But the other side is a huge research area

Gathering data from both longitudinal studies and from many types of cultures provide data on neurochemical effects on neural function.
Overall brain states (mental illness, disease) are believed to originate in the neurochemistry (research in depression and schizophrenia suggests this)
But there's a long way from neural to whole brain level (and we don't understand how one really gives rise to the other).

Concluding this section

Neuroinformatics

in the shape of data archiving, data analysis service provision, XML-based data and model description

Has a lot to offer experimental neuroscience, computational neuroscience, and clinical neuroscience, (but perhaps a little less to offer cognitive neuroscience)

Though connecting behavioural and neural data may be relevant here

Why has Neuroinformatics taken so long to get going?

- **Metadata difficulties**
  - **both for data and for models: a major international task**
- **Sociological difficulties**
  - **getting the different communities to talk to each other**
  - **difficulties with ownership of data**
  - **issues of security.**

Section 2: What's happening in Neuroinformatics?

The OECD Neuroinformatics Working Group and the International Neuroinformatics Coordinating Facility (INCF)

The CARMEN project.

The International Neuroinformatics Coordinating Facility (INCF)

OECD Neuroinformatics Working Group identified the need to work cooperatively in order to achieve major advances

This led to the setting up of the International Neuroinformatics Coordinating Facility (INCF)

See http://www.incf.org/

This has Czech republic, Finland, France, Germany, Italy, Japan, Norway, Sweden, Switzerland, and the USA as current members.

Other countries (specifically the UK) are considering joining.

INCF Mission statement

The mission of the INCF is to coordinate and foster international activities in Neuroinformatics. The INCF will contribute to the development and maintenance of database and computational infrastructure and support mechanisms for neuroscience applications. This infrastructure will enable access to all freely accessible data and analysis resources for human brain research to the international research community. INCF will develop mechanisms for the seamless flow of information and knowledge between academia, private enterprises and the publication industry. The larger objective of INCF is to contribute to the development of scalable, portable, and extensible applications that can be used by neuroscience laboratories for furthering our knowledge of the human brain and its diseases.

(from the INCF website)

The CARMEN project

Code Analysis, Repository and Modelling for e-Neuroscience

CARMEN is a new UK research council funded project in Neuroinformatics.

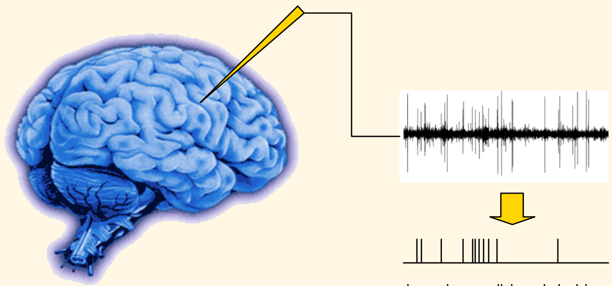
It aims to take advantage of the existing e-Science infrastructure to enable data archiving, secure data sharing, and configurable and extensible services for data analysis and manipulation.

PI: Prof Colin Ingram, Institute of Neuroscience, University Newcastle

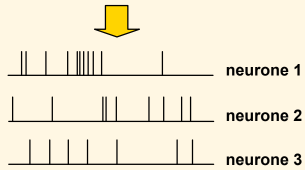
11 UK Universities, 19 Investigators, including experimentalists, modellers and spike train analysts.

CARMEN – Focus on Neural Activity

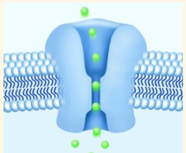
- **raw voltage signal data collected by patch-clamp and single & multi-electrode array recording**
- **optical recording, particularly the activity dynamics of large networks**



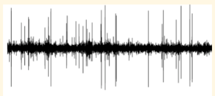
- **resolving the 'neural code' from the timing of action potential activity**



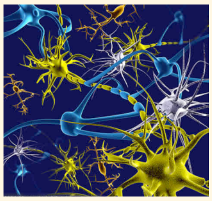
CARMEN - Scales of Integration



— determining ion channel contribution to the timing of action potentials



— resolving the 'neural code' from the timing of action potential activity



— examining integration within networks of differing dimensions

CARMEN Objectives

- To create a grid-enabled, real time 'virtual laboratory' environment for neurophysiological data
- To develop an extensible 'toolkit' for data manipulation extraction, analysis and modelling
- To provide a repository for archiving, sharing, integration and discovery of data
- To achieve wide community and commercial engagement in developing and using CARMEN

CARMEN is a 4 year project: if it is to last longer, it must become financially self-sufficient.

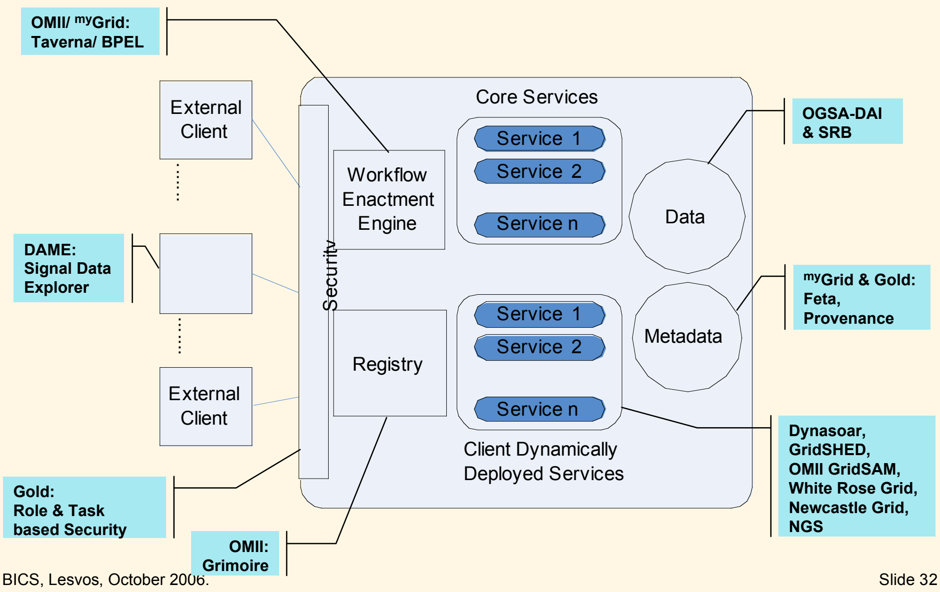
CARMEN Structure

Hub and Spoke Project

Hub:
A "CAIRN" repository for the storage and analysis of neuroscience data

Spokes:
A set of neuroscience projects that will produce data and analysis services for the hub, and use it to address real neuroscience questions

CARMEN Active Information Repository Node



Spokes: The Work Packages

WP1 Spike Detection & Sorting

WP2 Information Theoretic Analysis of Derived Signals

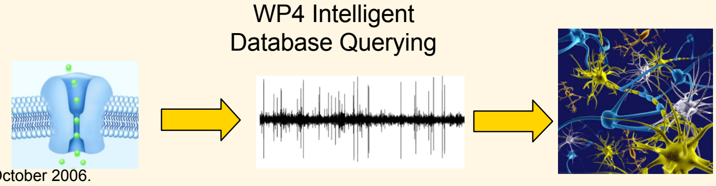
WP 3 Data-Driven Parameter Determination in Conductance-Based Models

Data Storage & Analysis

WP4 Measurement and Visualisation of Spike Synchronisation

WP5 Multilevel Analysis and Modelling in Networks

WP4 Intelligent Database Querying



CARMEN Consortium

Leadership & Infrastructure

ION The Institute of Neuroscience — UNIVERSITY OF NEWCASTLE UPON TYNE

Colin Ingram

north-east regional e-science centre — neresc

Paul Watson

UNIVERSITY OF STIRLING

White Rose university consortium — Universities of Leeds, Sheffield & York

THE UNIVERSITY of York

Leslie Smith **Jim Austin**

CARMEN Consortium

Work Packages

UNIVERSITY OF CAMBRIDGE

University of St Andrews

University of Leicester

MANCHESTER 1824
The University of Manchester

THE UNIVERSITY OF WARWICK

The University Of Sheffield

Imperial College London

CARMEN and the metadata problem

Metadata scheme required for

- **Experimental data description**
- **Experiment description**
- **Service description**
- **Model description**
  - **And possibly other tasks as well**

Re-use existing schemes

- **Don't create a new scheme**
- **But add elements if required, and try to get them adopted by BrainML or NeuroML as appropriate**

CARMEN has already forged links with Dan Gardner (BrainML), and with Shiro Usui (Neuroinformatics, Riken), both of whose groups are already involved in the metadata problem.

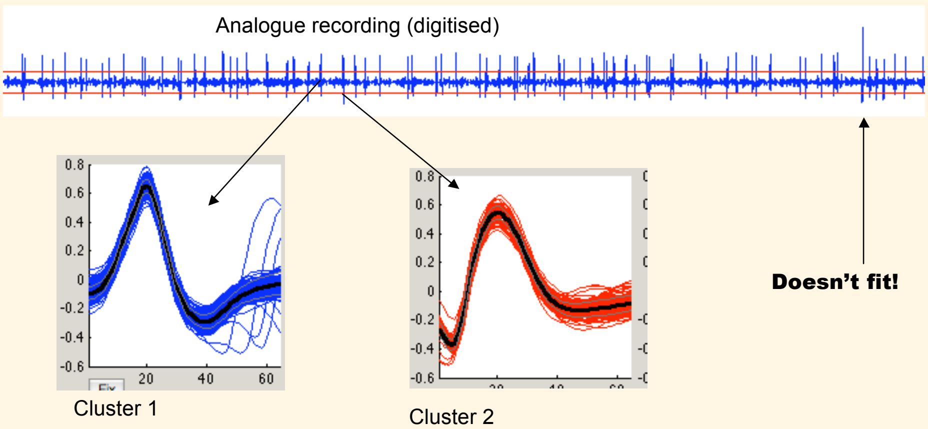
The CARMEN consortium hope to work in close collaboration with the (hopefully about to be created) UK INCF node.

Section 3: Neuroinformatics and spikes

Some current research topics

- **Spike detection and sorting**
- **Metrics on spike trains**
- **Detecting correlation neurally**

Spike Detection & Sorting (CARMEN WP1)



(Clustering using wave_clus (Quian-Quiroga))

Techniques for spike detection and sorting

Detection techniques:
- Simple thresholding
- Energy-based techniques

Sorting techniques
- Extract samples from segment around detected spike
- Reduce dimensionality
- Cluster many examples into a small number of clusters
- Label detected spikes
- Look at unassigned spikes: attempt to label as sum (collisions)

Easy if there's one type of spike, and good SNR
Can be difficult if SNR poor and multiple spike types
- Extracellular recording with MEAs

Experimental neuroscientists would like feedback while doing the experiment!

(In fact they often listen to spikes, relying on their auditory system to tell them when they have a good SNR.
- But they can't do spike sorting this way!)

CARMEN and spike detection and sorting

Idea is to provide many services

- Several different types of spike detection algorithms
- Several different types of spike sorting techniques
  - (including different types of data reduction, as well as different types of clustering)

Allow the user to test with a variety of techniques, and then choose the techniques they prefer

High speed links should allow immediate transfer of some datasets to Grid based systems

- Allow experimentalist to choose near-real-time detection and sorting for immediate feedback
  - To assist during the experiment
- Slower (and more effective) techniques for later analysis off-line.

Allow comparison of different techniques on a wide variety of data
- Which is best, and for what?

CARMEN and spike train analysis

CARMEN will store
- **Raw recorded data**
- **Detected and sorted spike train**

Beyond this level lies spike train analysis:
- **What do spike trains tell us about stimuli?**
- **What do spike trains encode, and how do they encode it?**

Many techniques have been used
- **Correlation based, Information theory based.**

CARMEN will apply these techniques to large volumes of data.

It will also examine new techniques.

Metrics for spike trains (with Thomas Wennekers, Manik Bhattarchee, Bartosz Telenczuk)

Metrics define the distance between two elements of a vector space.
Spike trains can be considered as elements of a metric space

- **But what metric to use?**

Spike trains are considered as sequences of spike times.

Euclidean Metrics
- **Turn spike train into a real-valued function, and measure the distance between them using a Euclidean metric.**
- **(e.g. represent each spike as a Dirac Delta function, and convolve with (e.g.) a Gaussian)**

Non-Euclidean metrics
- **Consider how to turn one spike train into another, costing each change made.**
- **Calculate the minimum cost over all possible ways of converting the spike trains**
- **(Victor, 1995)**

Clearly there are many families of metrics

What can we do with metrics?

Many different metrics exist
Need to investigate which are best for finding which types of similarities.
- **e.g. Correlations**
- **Spike trains from repeated stimuli**
- **Spike trains from related cell types.**

Metrics can be used to determine clusters of spike trains

Consider the neighbourhood of a spike train s_1:
That is,

$$N(s_1, \delta) = \{s : d(s_1, s) < \delta\}$$

By choosing δ appropriately, we can classify different sets of spike trains.

Clearly, the neighborhoods found will depend on the metric *d*, as well as δ.

Use metrics to help compare spike trains over time, and across experiments.

Analysing multiple spike trains for correlation (Bofill-y-petit and Murray, Richert)

Say we have N spike trains, n of which are correlated.

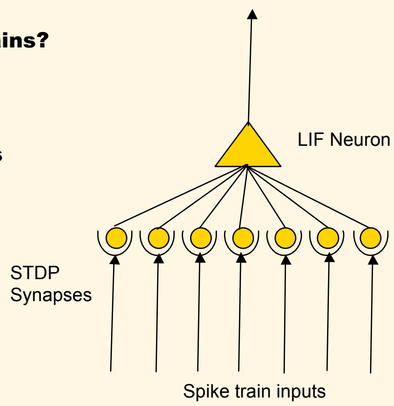
How might we find these n?

Direct correlation, pair wise?
Calculate distances between pairs of spike trains?

An alternative is to use a neural technique:

Apply the spike trains to randomly initialised excitatory synapses. Effect is that the weights of the correlated neurons rise, and those of the uncorrelated neurons decrease. Each synapse is an STDP synapse. When the neuron fires, those synapses responsible for the firing are strengthened, and those not (or less) so are penalised. Effect is that the weights from the correlated neurons increase, and those from the uncorrelated neurons decrease.

By inspecting the final weights, the correlated inputs can be found.



Summary and Conclusions

Neuroinformatics will have impact on a variety of areas of Neuroscience:

- **Experimental,**
- **Computational,**
  - **(both data, analysis and modelling)**
- **Cognitive, and**
- **Clinical**

The International Neuroinformatics Co-ordinating Facility (INCF) recognises the international dimension in this.

- **The new UK project, CARMEN, hope to advance Neuroinformatics to achieve its objectives**

Neuroinformatics represents a novel direction for Neuroscience: it will supply novel insights for Experimental, Computational, Cognitive and Clinical Neuroscience

- **through permitting re-use (and hence re-interpretation and integration) of datasets,**
- **through providing better data for model definition and evaluation**
- **through enabling better communication about models.**

Acknowledgements

EPSRC

The other investigators on the CARMEN project:
Jim Austin , Colin Ingram, Paul Watson , Stuart Baker, Roman Borisyuk, Stephen Eglen, Jianfeng Feng, Kevin Gurney, Tom Jackson, Marcus Kaiser, Phillip Lord, Stefano Panzeri, Rodrigo Quian Quiroga, Simon Schultz, Evelyne Sernagor, V. Anne Smith, Tom Smulders, Miles Whittington.

Thomas Wennekers, Micah Richert, Manik Bhattacharjee, Bartosz Telenczuk and the organisers of the Spike Train Analysis Workshop (particularly Stuart Baker (again!)), Newcastle, September 2006.

Nhamo Mtetwa